Indag. Mathem., N.S., 6 (2), 247–255 June 26, 1995

Monotonic functions and an inequality of Myerson on point distributions

by Reinhard Winkler*

Institut für Algebra und Diskrete Mathematik der TU Wien, Wiedner Hauptstrasse 8–10, A-1040 Wien, Austria

Communicated by Prof. R. Tijdeman at the meeting of May 30, 1994

ABSTRACT

In the paper 'Discrepancy and distance between sets' G. Myerson studied several notions of distances and discrepancies of point distributions on the unit interval. Among several results he proves an inequality between p-discrepancy ($p \geq 1$) and the 'distance' between n-element subsets of the unit interval. This paper contains a proof of his conjecture that this inequality holds in a stronger version. Furthermore it can be transferred to arbitrary probability distributions on the unit interval which are Borel measures. The results can be embedded into a topological context. The last section contains a corollary which is a further expansion of the inequality's domain of validity.

1. INTRODUCTION

In [M] G. Myerson considers finite subsets S and T of the unit interval $I = [0,1]$ having the same cardinality n, say

$$S = \{s_1 < \cdots < s_n\}, \qquad T = \{t_1 < \cdots < t_n\}.$$

By

$$|S - T| = \max\{|s_i - t_i| \mid i = 1, \ldots, n\}$$

a metric on the system of all n-element subsets is defined. For the set S one considers the induced distribution function

$$g_S(x) = \frac{1}{n}\#\{i = 1, \ldots, n \mid s_i \leq x\}.$$

* Institut für Algebra und Diskrete Mathematik der TU Wien, Kommission für Mathematik der Österreichischen Akademie der Wissenschaften.

For $p \geq 1$ the p-discrepancy $D_p(S)$ (cf. for instance the textbook [KN]) is defined as the L_p-distance $d_p(g_S, x)$ between g_S and the distribution function x of the continuous uniform distribution on I (i.e. the Lebesgue measure). Recall that for arbitrary $f, g \in L_p$ the L_p-distance $d_p(f, g)$ is defined by

$$d_p(f,g) = \left(\int_0^1 |f(x) - g(x)|^p dx \right)^{1/p}.$$

Myerson proves

$$|D_p(S)^p - D_p(T)^p| \leq c|S - T|$$

with some $c \leq 2$ and conjectures that the inequality holds for $c = 1$. For $c < 1$ counterexamples can be found.

If one defines for arbitrary nondecreasing maps (or, which is essentially the same, for arbitrary distribution functions of Borel probability measures on I) f and $g : I \to I$ the distance $d(f, g)$ by

$$\begin{aligned} &d(f,g) \\ &\quad = \inf\{\varepsilon > 0 \,|\, f(x+\varepsilon) \geq g(x) \text{ and } g(x+\varepsilon) \geq f(x) \text{ for all } x \in [0, 1-\varepsilon]\} \end{aligned}$$

one observes $d(g_S, g_T) = |S - T|$. Hence Myerson's conjecture follows from

Theorem 1.

$$|d_p(f,v)^p - d_p(g,v)^p| \leq d(f,g)$$

for all nondecreasing $f, g, v : I \to I$. *In particular* (*for* $v = g$) *this implies*

$$d_p(f,g)^p \leq d(f,g)$$

and (*for* $v \equiv 0$)

$$\left| \int_I (f(x)^p - g(x)^p) dx \right| \leq d(f,g).$$

After several preparations in Section 2 the proof of Theorem 1 will be given in Section 3 and is the main object of this paper. Section 4 presents a corollary of Theorem 1 concerning an arbitrary (not necessarily monotonic) measurable $v : I \to I$.

2. SEVERAL PREPARATIONS FOR THE PROOF OF THEOREM 1

Before we start with the proof of Theorem 1 we present a few remarks and notations:

d and d_p are pseudometrics on the set of all monotonic nondecreasing functions $f : I \to I$. For such an f the right and left side limits

$$f(x^+) = \inf\{f(y) \,|\, y > x\} \quad \text{for } x < 1$$

and

$$f(x^-) = \sup\{f(y) \,|\, y < x\} \quad \text{for } x > 0$$

exist. It is easy to see that $d(f,g)=0$ iff $d_p(f,g)=0$ iff $f(x^+)=g(x^+)$ for all $x\in[0,1)$ iff $f(x^-)=g(x^-)$ for all $x\in(0,1]$ iff f and g differ at most on a countable set of common discontinuities. In this case we write $f\sim g$ and may identify f and g. In other words: We may consider $f\in\mathcal{M}$ to be defined by all the values $f(x^+)$, $x\in[0,1)$ where

$$\mathcal{M}=\{f:I\to I\mid f(x)\le f(y)\text{ for all }x\le y\in I\}/_\sim$$

is the factor set with respect to the equivalence $\sim$. Correspondingly, d and d_p are metrics on $\mathcal{M}$. Sometimes it is convenient to choose the unique representative of an equivalence class which satisfies $f(1)=1$ and $f(x)=f(x^+)$, i.e. which is continuous from the right. Note that this f is the distribution function of a measure. Thus we have a bijective correspondence between all $f\in\mathcal{M}$ and all probability measures μ on I defined on the σ-algebra of Borel sets if we put $\mu([0,x])=f(x^+)$.
Several times we shall use that for all $a,b\in\mathbf{R}$ the function

$$\sigma_{a,b}(y)=|a-y|^p-|b-y|^p$$

is monotonically nondecreasing if $a<b$ and nonincreasing if $b<a$. This follows from the convexity of the function $x\mapsto|x|^p$ for $p\ge 1$. For us the case $a=f(x)$, $b=g(x)$ will be of interest.

Another observation simplifying the proof of Theorem 1 is that it suffices to consider $f,g,v\in\mathcal{S}\subseteq\mathcal{M}$ where $\mathcal{S}$ is the set of monotonic step functions, more precise: $f\in\mathcal{S}$ iff $f\in\mathcal{M}$ and there is a finite set

$$\{0=x_0<x_1<\cdots<x_n=1\}\subseteq I$$

such that $f\equiv y_i$ is a constant on each $J_i=(x_{i-1},x_i)$, $i=1,\ldots,n$:

Lemma. *Theorem* 1 *follows if*

$$|d_p(f,v)^p-d_p(g,v)^p|\le d(f,g)$$

holds for all $f,g,v\in\mathcal{S}$.

Proof. First we prove claim 1: Let $f\in\mathcal{M}$ and $\varepsilon>0$. Then there is an $s\in\mathcal{S}$ with $d_p(f,s)<\varepsilon$ and $d(f,s)<\varepsilon$.
Proof of claim 1: First choose an integer $n>1/\varepsilon$ and define

$$y_i=\inf\left\{x\in I\;\middle|\;f(x)\ge\frac{i}{n}\text{ or }x=1\right\},\quad i=0,\ldots,n.$$

Then put $z_i=i/n$, $i=0,\ldots,n$, and let

$$X=\{x_0=0<x_1<\cdots<x_k=1\}$$

be the set of all y_i and z_i. Now define s by $s\equiv f(x_{i-1}^+)$ on $J_i=(x_{i-1},x_i)$, $i=1,\ldots,n$. Then, by construction of the y_i, we have $|f(x)-s(x)|\le 1/n$ for (almost) all $x\in I$, hence $d_p(f,s)\le 1/n<\varepsilon$, and by the choice of the z_i we also have $d(f,s)\le 1/n<\varepsilon$.

To finish the proof of the lemma use claim 1 to get $s_f, s_g, s_v \in \mathcal{S}$ such that $d_p(f, s_f)$, $d_p(g, s_g)$, $d_p(v, s_v)$, $d(f, s_f)$ and $d(g, s_g)$ are sufficiently small and derive

$$|d_p(f,v)^p - d_p(g,v)^p| \le |d_p(s_f,s_v)^p - d_p(s_g,s_v)^p| \le d(s_f,s_g) + \frac{\varepsilon}{2}$$
$$\le d(f,g) + \varepsilon.$$

Since $\varepsilon > 0$ can be taken arbitrarily this proves the lemma. □

3. PROOF OF THEOREM 1

By the lemma of Section 2 everything we have to prove is $|F(f,g,v)| \le d(f,g)$ for arbitrary $f, g, v \in \mathcal{S}$ where

$$F(f,g,v) = F_I(f,g,v) = d_p(f,v)^p - d_p(g,v)^p,$$

$$F_J(f,g,v) = \int_J \psi(f,g,v,x)\,dx \quad \text{and}$$

$$\psi(f,g,v,x) = |f(x) - v(x)|^p - |g(x) - v(x)|^p.$$

The idea of the proof of Theorem 1 is, starting with $(f_0, g_0, v_0) = (f, g, v)$, to construct a finite chain (f_i, g_i, v_i), $i = 0, 1, 2, 3, 4, 5$, of triplets such that

$$F(f_i, g_i, v_i) \le F(f_{i+1}, g_{i+1}, v_{i+1}) \quad \text{and} \quad d(f_{i+1}, g_{i+1}) \le d(f_i, g_i).$$

We will finish with $v_5 \equiv 0$, $f_5 \equiv 1$ and $g_5 = \chi_{[\varepsilon, 1]}$ such that

$$F(f,g,v) \le F(f_5, g_5, v_5) = \varepsilon = d(f_5, g_5) \le d(f,g).$$

Changing the roles of f and g one also gets

$$-F(f,g,v) = F(g,f,v) \le d(g,f) = d(f,g),$$

yielding Theorem 1.

For $f, g, v \in \mathcal{M}$ we call an interval $J \subseteq I$ (f,g,v)-regular if either $f(x) \le g(x) \le v(x)$ for all $x \in J$ or $v(x) \le g(x) \le f(x)$ for all $x \in J$. Similarly call J (f,g)-regular if either $f(x) \le g(x)$ or $g(x) \le f(x)$ for all $x \in J$.

$$J_i = (x_{i-1}, x_i), \quad i = 1, \dots, n, \qquad (0 = x_0 < x_1 < \cdots < x_n = 1)$$

is called an (f,g,v)- resp. (f,g)-regular decomposition if each J_i is (f,g,v)- resp. (f,g)-regular but no (x_{i-1}, x_{i+1}). For $f, g \in \mathcal{S}$ there always exists an (f,g)-regular decomposition. A property of regular decompositions which will be crucial in the proof is that

$$\sup\{f(x), g(x), v(x) \mid x \in J_{i-1}\} \le \inf\{f(x), g(x), v(x) \mid x \in J_i\}$$

resp.

$$\sup\{f(x), g(x) \mid x \in J_{i-1}\} \le \inf\{f(x), g(x) \mid x \in J_i\}$$

for $i = 2, \dots, n$. This guarantees that within each J we may change the values of the functions without destroying monotonicity between different regular intervals as long as the bounds sup and inf in J are valid. Similarly $d(f,g)$ with respect to the whole interval I cannot increase if it does not with respect to J.

To describe the subsequent constructions it is convenient to define the following properties P_j for a triplet (f, g, v):

- P_0: $f, g, v \in \mathcal{S}$.
- P_1: There exists an (f, g)-regular decomposition J_i, $i = 1, \ldots, n$, such that v is a constant on each J_i.
- P_2: For every $x \in I$ either $v(x) \leq g(x) \leq f(x)$ or $f(x) \leq g(x) \leq v(x)$. Together with P_0 this guarantees the existence of an (f, g, v)-regular decomposition.
- P_3: $v(x) \leq g(x) \leq f(x)$ for all $x \in I$.
- P_4: $v \equiv 0$ and $g(x) \leq f(x)$ for all $x \in I$.
- P_5: $v \equiv 0, f \equiv 1$ and $g = \chi_{[\varepsilon, 1]}$.

For $j = 0, 1, 2, 3, 4$ we shall show that for any given triplet (f, g, v) with P_0 and P_j there exists a triplet (f', g', v') with

- (a) P_0,
- (b) P_{j+1},
- (c) $F(f, g, v) \leq F(f', g', v')$ and
- (d) $d(f', g') \leq d(f, g)$.

This will give rise to a chain as announced above.

$j = 0$: Put $f' = f$, $g' = g$ and define v' as follows: fix any (f, g)-regular decomposition. For every $J = (a, b)$ of this decomposition define $v' \equiv v(a^+)$ on J if $g \leq f$ on J and $v' \equiv v(b^-)$ otherwise. For (f', g', v') (a), (b) and (d) are trivial. To derive (c) just use the monotonicity of $\sigma_{f(x), g(x)}$ to derive

$$\psi(f', g', v', x) = \psi(f, g, v', x) \geq \psi(f, g, v, x)$$

for every $x \in I$.

$j = 1$: Fix an (f, g)-regular decomposition as guaranteed by P_1. Put $f' = f$ and $v' = v$. To define g' distinguish, for each J of the decomposition, two cases:

First case: $f \leq g$ on J. $J = J_1 \cup J_2 \cup J_3$ with $f \leq g \leq v$ on J_1, $f \leq v < g$ on J_2 and $v < f \leq g$ on J_3. Put $g' = g$ on J_1, $g' = v$ on J_2 and $g' = f$ on J_3.

Second case: $g \leq f$ on J. $J = J_1 \cup J_2 \cup J_3$ with $g \leq f \leq v$ on J_1, $g \leq v < f$ on J_2 and $v < g \leq f$ on J_3. Put $g' = f$ on J_1, $g' = v$ on J_2 and $g' = g$ on J_3.

It may be left to the reader to check that (a), (b), (c) and (d) hold.

$j = 2$: Consider an (f, g, v)-regular decomposition as in P_2 and apply to each corresponding (f, g, v)-regular $J = (a, b)$ the following construction. If $v \leq g \leq f$ on J put $f'|_J = f|_J$, $g'|_J = g|_J$ and $v'|_J = v|_J$ for the restrictions on J. Otherwise consider the minimal rectangle $R = J \times [c, d]$ such that the graphs of $f|_J$, $g|_J$ and $v|_J$ are contained in R. Execute a half rotation (i.e. angle $= \pi$) of R around its centre and take the images of the graphs as the new graphs of $f'|_J$, $g'|_J$ and $v'|_J$. The property of an (f, g, v)-regular decomposition mentioned above guarantees P_0 for (f', g', v'), i.e. (a). Note that the values of F and d do not change: $F(f', g', v') = F(f, g, v)$ and $d(f', g') = d(f, g)$, hence (c) and (d). Finally also (b), i.e. $v' \leq g' \leq f'$, is obvious.

$j = 3$: It is trivial that (f', g', v') with $f' = f$, $g' = g$ and $v' \equiv 0$ satisfies (a), (b) and (d). For (c) use the monotonicity of $\sigma_{f(x), g(x)}$.

$j = 4$: Let $\varepsilon = d(f, g)$, $f' \equiv 1$, $v' = v \equiv 0$ and g' be the characteristic function of the interval $[\varepsilon, 1]$ then (a), (b) and (d) are trivial. It remains to prove (c) which means

$$F(f, g, 0) = \int_I (f(x)^p - g(x)^p) dx \leq \varepsilon.$$

For every $y \in I$ we introduce the auxiliary functions

$$f_y(x) = \min(f(x), y) \quad \text{and} \quad g_y(x) = \min(g(x), y).$$

Furthermore we define

$$G(y) = \int_I (f_y(x)^p - g_y(x)^p) dx \quad \text{and} \quad G_0(y) = \varepsilon y^p.$$

Let be all values of f and g among the numbers

$$0 = y_0 < y_1 < \cdots < y_n = 1.$$

We shall show

$$G(y_i) - G(y_{i-1}) \leq \varepsilon(y_i^p - y_{i-1}^p) \quad \text{for } i = 1, \ldots, n.$$

By $G(y_0) = G(0) = 0 = y_0^p$ then the relation

$$F(f, g, 0) = G(1) - G(0) = \sum_{i=1}^n (G(y_i) - G(y_{i-1}))$$
$$\leq \varepsilon \sum_{i=1}^n (y_i^p - y_{i-1}^p) = \varepsilon(y_n^p - y_0^p) = \varepsilon$$

will follow.
The sets

$$A = \{x \in I \mid g(x) \leq f(x) \leq y_{i-1}\},$$
$$B = \{x \in I \mid g(x) \leq y_{i-1}, y_i \leq f(x)\} \quad \text{and}$$
$$C = \{x \in I \mid y_i \leq g(x) \leq f(x)\}$$

form a partition of I. Hence

$$G(y_i) = \int_A (f(x)^p - g(x)^p) dx + \int_B (y_i^p - g(x)^p) dx \quad \text{and}$$

$$G(y_{i-1}) = \int_A (f(x)^p - g(x)^p) dx + \int_B (y_{i-1}^p - g(x)^p) dx,$$

therefore

$$G(y_i) - G(y_{i-1}) = \int_B (y_i^p - y_{i-1}^p) dx = \lambda(B)(y_i^p - y_{i-1}^p).$$

$d(f, g) = \varepsilon$ implies $\lambda(B) \leq \varepsilon$, thus

$$G(y_i) - G(y_{i-1}) \leq \varepsilon(y_i^p - y_{i-1}^p). \qquad \square$$

Remark. The inequality $d_p(f,g)^p \leq d(f,g)$ expresses that the topology $\mathcal{O}_d$ induced by d is stronger than $\mathcal{O}_p$, the topology induced by d_p. In symbols: $\mathcal{O}_p \subseteq \mathcal{O}_d$. One may ask for the relations among other topologies on $\mathcal{M}$:

Let $\mathcal{O}$ denote the quotient topology of the compact Tychonoff topology (describing pointwise convergence) after the identification via $\sim$, cf. Section 2. Then $\mathcal{O}$ is compact and Hausdorff.

Furthermore one may consider $\mathcal{O}_u$, the topology of uniform convergence (of all right side limits $f(x^+)$) and $\mathcal{O}_w$, the weak topology, which is the weakest topology on $\mathcal{M}$ such that for all continuous $v : I \to I$ the map

$$\Phi : \mathcal{M} \to \mathbf{R}, \qquad f \mapsto \int_I v(x)\, df(x^+)$$

is continuous. Here the correspondence between $f \in \mathcal{M}$ and Borel probability measures on I is used (cf. Section 2). Using uniform continuity of v and a standard approximation argument one sees that convergence w.r.t. $\mathcal{O}$ implies weak convergence, i.e. $\mathcal{O}_w \subseteq \mathcal{O}$. Since $\mathcal{O}$ is compact and Hausdorff and $\mathcal{O}_w$ is Hausdorff too, this implies $\mathcal{O}_w = \mathcal{O}$. Similarly one proves $\mathcal{O}_p = \mathcal{O}$. (That there is a metric d' inducing $\mathcal{O}_w$ follows from the fact that I has a countable topological base, cf. [B], 46.4, p. 233. Since $\mathcal{O}_w = \mathcal{O} = \mathcal{O}_p$, every d_p, $p \geq 1$, can be taken as d'.)

The question whether there hold further inclusions can be answered negatively by the following examples:

The sequence $f_n \equiv 1/n$ converges to $f \equiv 0$ in the sense of $\mathcal{O}$ and $\mathcal{O}_u$ but not in the sense of $\mathcal{O}_d$.

The sequence $f_n = \chi_{[1/n,1]}$ (characteristic function of the interval $[1/n, 1]$) converges to $f \equiv 1$ in the sense of $\mathcal{O}$ and $\mathcal{O}_d$ but not of $\mathcal{O}_u$.
Thus we may summarize

$$\mathcal{O} = \mathcal{O}_w = \mathcal{O}_p \begin{array}{l} \subset \mathcal{O}_u \\ \subset \mathcal{O}_d \end{array}$$

where the inclusions $\subset$ are strict and neither $\mathcal{O}_u \subseteq \mathcal{O}_d$ nor $\mathcal{O}_d \subseteq \mathcal{O}_u$.

4. A COROLLARY OF THEOREM 1

Theorem 2. *Let $f, g \in \mathcal{M}$ and $v : I \to I$ be an arbitrary measurable selfmap of I. Then*

$$|d_p(f,v)^p - d_p(g,v)^p| \leq 2d(f,g).$$

Proof. W.l.o.g. (cf. the discussion in Section 2) let f and g be continuous from the right. Define the sets

$$A = \{x \in I \mid g(x) \leq f(x)\}, \qquad B = \{x \in I \mid f(x) \leq g(x)\}$$

and, as in Section 3, the function

$$\psi(x) = |f(x) - v(x)|^p - |g(x) - v(x)|^p.$$

Then

$$d_p(f,v)^p - d_p(g,v)^p = \int_A \psi(x)dx + \int_B \psi(x)dx.$$

To estimate the first term we use, as in Section 3, the monotonicity of the function $\sigma_{f(x),g(x)}$ to derive

$$\int_A \psi(x)dx \leq \int_A (f(x)^p - g(x)^p)dx.$$

Consider the restricted Lebesgue measures

$$\lambda_A(M) = \lambda(M \cap A), \qquad \lambda_J(M) = \lambda(M \cap J)$$

with $J = [0, \lambda(A)]$ and the map

$$\tau : J \to I, \qquad \tau(x) = \inf\{y \in A \mid \lambda_A([0,y]) \geq x\}.$$

Since f and g are continuous from the right, A is closed under limits from the right, hence $\tau(x) \in A$ for all $x \in J$. Thus, by definition, $\tau : J \to A$ is measure preserving with respect to λ_J and λ_A. Hence

$$\begin{aligned}\int_A (f(x)^p - g(x)^p)dx &= \int_A (f(x)^p - g(x)^p)d\lambda_A(x)\\ &= \int_J (f(\tau(x))^p - g(\tau(x))^p)d\lambda_J(x)\\ &= \int_I (f_A(x)^p - g_A(x)^p)dx\end{aligned}$$

if we define $f_A(x) = f(\tau(x))$, $g_A(x) = g(\tau(x))$ for $x \in J$ and $f_A(x) = g_A(x) = 1$ for $x > \lambda(A)$. τ satisfies $|\tau(y) - \tau(x)| \geq |y - x|$ which implies $d(f_A, g_A) \leq d(f,g)$. Since $f_A, g_A \in \mathcal{M}$ we may apply Theorem 1 to summarize

$$\begin{aligned}\int_A \psi(x)dx &\leq \int_A (f(x)^p - g(x)^p)dx = \int_I (f_A(x)^p - g_A(x)^p)dx\\ &= d_p(f_A,0)^p - d_p(g_A,0)^p \leq d(f_A,g_A) \leq d(f,g).\end{aligned}$$

Similarly one defines f_B and g_B to get

$$\begin{aligned}\int_B \psi(x)dx &\leq \int_B |f(x)-1|^p - |g(x)-1|^p dx\\ &= d_p(f_B,1)^p - d_p(g_B,1)^p \leq d(f_B,g_B) \leq d(f,g).\end{aligned}$$

Thus

$$d_p(f,g)^p - d_p(g,v)^p \leq 2d(f,g)$$

and, by symmetry,

$$-(d_p(f,v)^p - d_p(g,v)^p) = d_p(g,v)^p - d_p(f,v)^p \leq 2d(g,f) = 2d(f,g),$$

proving

$$|d_p(f,v)^p - d_p(g,v)^p| \leq 2d(f,g). \qquad \square$$

Theorem 2 is best possible in the following sense: for every $c < 2$ there are $f, g \in \mathcal{M}$, $v : I \to I$ measurable and $p \geq 1$ such that

$$|d_p(f,v)^p - d_p(g,v)^p| > cd(f,g).$$

As an example define for any fixed $\varepsilon > 0$ f, g and v in the following way: if $0 \leq x \leq \varepsilon$ then $f(x) = 0, g(x) = \frac{1}{2}$ and $v(x) = 1$. If $\varepsilon < x \leq 2\varepsilon$ then $f(x) = 1, g(x) = \frac{1}{2}$ and $v(x) = 0$. If $x > 2\varepsilon$ then $f(x) = g(x) = v(x) = 1$. Then indeed

$$\begin{aligned}|d_p(f,v)^p - d_p(g,v)^p| &= 2\varepsilon - 2\varepsilon\left(\frac{1}{2}\right)^p = 2\left(1 - \left(\frac{1}{2}\right)^p\right)\varepsilon > c\varepsilon \\ &= cd(f,g)\end{aligned}$$

for sufficiently large p.

REFERENCES

[B] Bauer, H. – Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. Walter de Gruyter, Berlin–New York (third edition), (1978).

[KN] Kuipers, L. and H. Niederreiter – Uniform Distribution of Sequences. Wiley, New York (1974).

[M] Myerson, G. – Discrepancy and distance between sets. Indag. Mathem., N.S. 3 (2), 193–201 (1992).